

# SEQUENCE LISTING

## (1) GENERAL INFORMATION:

- (i) APPLICANT: Tsien et al., Roger Y. et al.
- (ii) TITLE OF INVENTION: LONG WAVELENGTH MUTANT FLUORESCENT PROTEINS
- (iii) NUMBER OF SEQUENCES: 4
- (iv) CORRESPONDENCE ADDRESS:
  - (A) ADDRESSEE: Fish & Richardson P.C.
  - (B) STREET: 4225 Executive Square, Suite 1400
  - (C) CITY: La Jolla
  - (D) STATE: CA
  - (E) COUNTRY: USA
  - (F) ZIP: 92037
- (v) COMPUTER READABLE FORM:
  - (A) MEDIUM TYPE: Floppy disk
  - (B) COMPUTER: IBM PC compatible
  - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
  - (D) SOFTWARE: PatentIn Release #1.0, Version #1.30
- (vi) CURRENT APPLICATION DATA:
  - (A) APPLICATION NUMBER: UNKNOWN
  - (B) FILING DATE: 15-AUG-1997
  - (C) CLASSIFICATION:
- (vii) ATTORNEY/AGENT INFORMATION:
  - (A) NAME: Haile, Lisa A.
  - (B) REGISTRATION NUMBER: 38,347
  - (C) REFERENCE/DOCKET NUMBER: 07257/056001
- (ix) TELECOMMUNICATION INFORMATION:
  - (A) TELEPHONE: 619/678-5070
  - (B) TELEFAX: 619/678-5099

## (2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 716 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

### (ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..714

### (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

ATG AGT AAA GGA GAA CTT TTC ACT GCA GTT GTC CCA ATT CTT GTT	48
Met Ser Lys Gly Glu Leu Phe Thr Ala Val Pro Ile Leu Val	
1 5 10 15	
GAA TTA GAT GGT CAT GTT AAT GGG CAC AAA TTT TCT GTC AGT GGA GAG	96
Glu Leu Asp Gly Asp Val Asn Gly His Lys Phe Ser Val Ser Gly Glu	
20 25 30	
GGT GAA GGT GAT GTA ACA TAC GGA AAA CTT ACC CTT AAA TTT ATT TGC	144
Gly Glu Gly Asp Val Thr Tyr Gly Lys Leu Thr Leu Lys Phe Ile Cys	
35 40 45	
ACT ACT GGA AFA CTA CCT GTT CCA TGG CCA ACA CTT GTC ACT ACT TTC	192
Thr Thr Gly Lys Leu Pro Val Pro Thr Pro Thr Leu Val Thr Thr Phe	
50 55 60	
TCT TAT GGT GIT CAA TGC TTT TCA AGA TAC CCA GAT CAT ATG AAA CGG	240
Ser Tyr Gly Val Gln Cys Phe Ser Arg Tyr Pro Asp His Met Lys Arg	
65 70 75 80	

CAT GAC TTT TTC AAG AGT GCC ATG CCC GAA GGT TAT GTA CAG CAA AGA 288  
 His Asp Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Val Gln Gln Arg  
 85 90 95  
 ACT ATA TTT TTC AAA GAT GAC GGG AAC TAC AAG ACA CGT GCT GAA GTC 336  
 Thr Ile Phe Phe Lys Asp Asp Gly Asn Tyr Lys Thr Arg Ala Glu Val  
 100 105 110  
 AAG TTT GAA GGT GAT ACC CTT GTT AAT AGA ATC GAG TTA AAA GGT ATT 384  
 Lys Phe Glu Gly Asp Thr Leu Val Asn Arg Ile Glu Leu Lys Gly Ile  
 115 120 125  
 GAT TTT AAA GAA GAT GGA AAC ATT CTT GGA CAT AAA TTG GAA TAC AAC 432  
 Asp Phe Lys Glu Asp Gly Asn Ile Leu Gly His Lys Leu Glu Tyr Asn  
 130 135 140  
 TAT AAC TCA CAC AAT GTA TAC ATC ATG GCA GAC AAA CAA AAG AAT GGA 480  
 Ile Asn Ser His Asn Val Tyr Ile Met Ala Asp Lys Gln Lys Asn Gly  
 145 150 155 160  
 ATC AAA GTT AAC TTC AAA ATT AGA CAC AAC ATT GAA GAT GGA AGC GTT 528  
 Ile Lys Val Asn Phe Lys Ile Arg His Asn Ile Glu Asp Gly Ser Val  
 165 170 175  
 GAA CTA GCA GAC TAT TAT CAA CAA AAT ACT CCA ATT CTC GAT GGC CCT 576  
 Gln Leu Ala Asp Tyr Tyr Gln Gln Asn Thr Pro Ile Leu Asp Gly Pro  
 180 185 190  
 GTC CTT TTA CCA GAC AAC CAT TAC CTG TCC ACA CAA TCT GCC CTT TCG 624  
 Val Leu Leu Pro Asp Asn His Tyr Leu Ser Thr Gln Ser Ala Leu Ser  
 195 200 205  
 AAA GAT CCC AAC GAA AAG AGA GAC CAC ATG GTC CTT CTT GAG TTT GTA 672  
 Lys Asp Pro Asn Glu Lys Arg Asp His Met Val Leu Leu Glu Phe Val  
 210 215 220  
 ACA GCT GCT GGG ATT ACA CAT GGC ATG GAT GAA CTA TAC AAA 714  
 Thr Ala Ala Gly Ile Thr His Gly Met Asp Glu Leu Tyr Lys  
 225 230 235  
 TTA 716

(1) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 238 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Ser Lys Gly Glu Glu Leu Phe Thr Ala Val Val Pro Ile Leu Val  
 1 5 10 15  
 Glu Leu Asp Gly Asp Val Asn Gly His Lys Phe Ser Val Ser Gly Glu  
 20 25 30  
 Gly Glu Gly Asp Val Thr Tyr Gly Lys Leu Thr Leu Lys Phe Ile Cys  
 35 40 45  
 Thr Thr Gly Lys Leu Pro Val Pro Trp Pro Thr Leu Val Thr Thr Phe  
 50 55 60  
 Ser Tyr Gly Val Gln Cys Phe Ser Arg Tyr Pro Asp His Met Lys Arg  
 65 70 75 80  
 His Asp Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Val Gln Gln Arg  
 85 90 95  
 Thr Ile Phe Phe Lys Asp Asp Gly Asn Tyr Lys Thr Arg Ala Glu Val  
 100 105 110  
 Lys Phe Glu Gly Asp Thr Leu Val Asn Arg Ile Glu Leu Lys Gly Ile  
 115 120 125  
 Asp Phe Lys Glu Asp Gly Asn Ile Leu Gly His Lys Leu Glu Tyr Asn  
 130 135 140

Tyr Asn Ser His Asn Val Tyr Ile Met Ala Asp Lys Gln Lys Asn Gly  
 145 150 155 160  
 Ile Lys Val Asn Phe Lys Ile Arg His Asn Ile Glu Asp Gly Ser Val  
 165 170 175  
 Gln Leu Ala Asp Tyr Tyr Gln Gln Asn Thr Pro Ile Leu Asp Gly Pro  
 180 185 190  
 Val Leu Leu Pro Asp Asn His Tyr Leu Ser Thr Gln Ser Ala Leu Ser  
 195 200 205  
 Lys Asp Pro Asn Glu Lys Arg Asp His Met Val Leu Leu Glu Phe Val  
 210 215 220  
 Thr Ala Ala Gly Ile Thr His Gly Met Asp Glu Leu Tyr Lys  
 225 230 235

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 720 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ix) FEATURE:  
 (A) NAME/KEY: CDS  
 (B) LOCATION: 1..720

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

ATG GTG AGC AAG GGC GAG GAG CTG TTC ACC GGG GTG GTG CCC ATC CTG 48  
 Met Val Ser Lys Gly Glu Glu Leu Phe Thr Gly Val Val Pro Ile Leu  
 240 245 250  
 GGC GAG CTG CAC GGC GAC GTA AAC GGC CAC AAG TTC AGC GTG TCC GGC 96  
 Val Glu Leu Asp Gly Asp Val Asn Gly His Lys Phe Ser Val Ser Gly  
 255 260 265 270  
 GAG GGC GAG GGC GAT GCC ACC TAC GGC AAG CTG ACC CTG AAG TTC ATC 144  
 Glu Gly Glu Gly Asp Ala Thr Tyr Gly Lys Leu Thr Leu Lys Phe Ile  
 275 280 285  
 TGC ACC ACC GGC AAG CTG CCC GTG CCC TGG CCC ACC CTC GTG ACC ACC 192  
 Cys Thr Thr Gly Lys Leu Pro Val Pro Trp Pro Thr Leu Val Thr Thr  
 290 295 300  
 TTC GGC TAC GGC GTG CAG TGC TTC GCC CGC TAC CCC GAC CAC ATG AAG 240  
 Phe Gly Tyr Gly Val Gln Cys Phe Ala Arg Tyr Pro Asp His Met Lys  
 305 310 315  
 CAG CAG GAC TTC TTC AAG TCC GCC ATG CCC GAA GGC TAC GTC CAG GAG 288  
 Gln Gln Asp Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Val Gln Glu  
 320 325 330  
 CGC ACC ATC TTC TTC AAG GAC GAC GGC AAC TAC AAG ACC CGC GCC GAG 336  
 Arg Thr Ile Phe Phe Lys Asp Asp Gly Asn Tyr Lys Thr Arg Ala Glu  
 335 340 345 350  
 GTG AAG TTC GAG GGC GAC ACC CTG GTG AAC CGC ATC GAG CTG AAG GGC 384  
 Val Lys Phe Glu Gly Asp Thr Leu Val Asn Arg Ile Glu Leu Lys Gly  
 355 360 365  
 ATC GAC TTC AAG GAC GAC GGC AAC ATC CTG GGG CAC AAG CTG GAG TAC 432  
 Ile Asp Phe Lys Asp Asp Gly Asn Ile Leu Gly His Lys Leu Glu Tyr  
 370 375 380  
 AAC TAC AAC AGC CAC AAC GTC TAT ATC ATG GCC GAC AAG CAG AAG AAC 480  
 Asn Tyr Asn Ser His Asn Val Tyr Ile Met Ala Asp Lys Gln Lys Asn  
 385 390 395  
 GGC ATC AAG GTG AAC TTC AAG ATC CGC CAC AAC ATC GAG GAC GGC AGC 528  
 Gly Ile Lys Val Asn Phe Lys Ile Arg His Asn Ile Glu Asp Gly Ser  
 400 405 410

GTG CAG CCC GCC GAC CAC TAC CAG CAG AAC ACC CCC ATC GGC GAC GGC 576  
Val Gln Pro Ala Asp His Tyr Gln Gln Asn Thr Pro Ile Gly Asp Gly  
415 420 425

CCC GTG CTG CTG CCC GAC AAC CAC TAC CTG AGC TAC CAG TCC GCC CTG 624  
Pro Val Leu Leu Pro Pro Asn His Tyr Leu Ser Tyr Gln Ser Ala Leu  
435 440 445

AGC AAA GAC CCC AAC GAG AAG CGC GAT CAC ATG GTC CTG CTG GAG TTC 672  
Ser Lys Asp Pro Asn Glu Lys Arg Asp His Met Val Leu Leu Glu Phe  
450 455 460

GTG ACC GCC GCC GGG ATC ACT CAC GGC ATG GAC GAG CTG TAC AAG TAA 720  
Val Thr Ala Ala Gly Ile Thr His Gly Met Asp Glu Leu Tyr Lys \*  
465 470 475

(2) INFORMATION FOR SEQ ID NO:4:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 240 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Met Val Ser Lys Gly Glu Glu Leu Phe Thr Gly Val Val Pro Ile Leu  
1 5 10 15  
Ala Glu Leu Asp Gly Asp Val Asn Gly His Lys Phe Ser Val Ser Gly  
20 25 30  
Glu Gly Glu Gly Asp Ala Thr Tyr Gly Lys Leu Thr Leu Lys Phe Ile  
35 40 45  
Cys Thr Thr Gly Lys Leu Pro Val Pro Trp Pro Thr Leu Val Thr Thr  
50 55 60  
Phe Gly Tyr Gly Val Gln Cys Phe Ala Arg Tyr Pro Asp His Met Lys  
65 70 75 80  
Gln Gln Asp Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Val Gln Glu  
85 90 95  
Arg Thr Ile Phe Phe Lys Asp Asp Gly Asn Tyr Lys Thr Arg Ala Glu  
100 105 110  
Val Lys Phe Glu Gly Asp Thr Leu Val Asn Arg Ile Glu Leu Lys Gly  
115 120 125  
Ile Asp Phe Lys Asp Asp Gly Asn Ile Leu Gly His Lys Leu Glu Tyr  
130 135 140  
Asn Tyr Asn Ser His Asn Val Tyr Ile Met Ala Asp Lys Gln Lys Asn  
145 150 155 160  
Gly Ile Lys Val Asn Phe Lys Ile Arg His Asn Ile Glu Asp Gly Ser  
165 170 175  
Val Gln Pro Ala Asp His Tyr Gln Gln Asn Thr Pro Ile Gly Asp Gly  
180 185 190  
Pro Val Leu Leu Pro Asp Asn His Tyr Leu Ser Tyr Gln Ser Ala Leu  
195 200 205  
Ser Lys Asp Pro Asn Glu Lys Arg Asp His Met Val Leu Leu Glu Phe  
210 215 220  
Val Thr Ala Ala Gly Ile Thr His Gly Met Asp Glu Leu Tyr Lys \*  
225 230 235 240